

Краткий русский вариант статьи (не соответствует полностью английскому варианту)  
А.М. Makarieva (2001) "Variance of protein heterozygosity in different species of mammals with respect to the number of loci studied", *Heredity*, 87(1), 41-51.  
<http://www.bioticregulation.ru/pubs/ab.php?id=hered&lang=ru>

## Экологический порог внутривидовой генетической изменчивости

млекопитающих

А. М. Макарьева

К настоящему времени накоплено большое количество данных о внутривидовой генетической изменчивости млекопитающих. Вопрос о том, является ли эта изменчивость генетическим шумом, стирающим информацию генома вида, или же она содержит потенциальные возможности приспособления вида к новым условиям обитания, широко дискутируется [1-5]. Если это генетический шум, то внутривидовая генетическая изменчивость естественных видов в естественных условиях их экологической ниши должна ограничиваться сверху стабилизирующим отбором и не превышать определенного порогового значения, которое естественно называть экологическим порогом генетической изменчивости для данного вида. Непрерывность мутагенеза должна приводить к тому, что в естественных условиях величина генетической изменчивости всех популяций должна увеличиваться вплоть до экологического порога, т.е. должна быть ограничена снизу. Однако наблюдения показывают, что наблюдаемая генетическая изменчивость у разных видов млекопитающих сильно варьирует, либо опускаясь ниже средней величины для всего класса вплоть до нулевых значений, либо существенно ее превосходя (рис. 1).

В настоящей работе показано, что наблюдаемая большая дисперсия распределения генетической изменчивости млекопитающих связана с используемым способом ее измерения. Естественная же ширина распределения оказывается малой, т.е. величина генетической изменчивости у различных видов млекопитающих примерно одинакова и близка к среднему значению для всего класса. Это позволяет

рассматривать последнее значение как экологический порог генетической изменчивости, общий для всего класса млекопитающих.

Генетическая изменчивость может быть количественно определена по величине гетерозиготности генов, кодирующих белки. Гены (генные локусы), присутствующие в популяции в виде двух или более вариантов (аллелей), называются полиморфными. Если в диплоидном (содержащем две копии) геноме одной особи присутствуют две различные или одинаковые аллели полиморфного гена, то соответственный генный локус у этой особи называется гетерозиготным или гомозиготным, соответственно. Гетерозиготность особи равна отношению числа гетерозиготных локусов к числу всех генных локусов генома.

В геноме млекопитающих содержится более 50 тысяч различных генов, кодирующих белки. Такое количество генов не может быть исследовано ни в одном современном эксперименте. Обычно исследуется около  $L = 20-30$  случайно выбранных генных локусов [3]. Затем для каждой особи определяется отношение наблюдаемого числа гетерозиготных локусов к общему числу исследованных локусов. Далее производится усреднение результата по всем особям популяции. Получаемая таким образом средняя гетерозиготность популяции  $H_L$  характеризует генетическую изменчивость выбранной группы  $L$  локусов, но не равна гетерозиготности  $H$  всей кодирующей белки части генома.

При случайном распределении полиморфных локусов по всему диплоидному геному числа их встречаемости (множественность)  $l_p$  в наборах из  $L$  локусов, выбираемых случайно в различных частях генома, должны быть распределены по Пуассону со средней множественностью  $\bar{l}_p$ , т.е. вероятность  $p(l_p)$  обнаружить  $l_p$  полиморфных локусов в заданной выборке из  $L$  локусов равна:

$$p(l_p) = e^{-\bar{l}_p} \frac{\bar{l}_p^{l_p}}{l_p!}$$

Для построения распределения числа полиморфных локусов в геноме данного вида необходима большая серия измерений, в которых заданное число локусов  $L$  выбирается случайно в различных частях генома.

Если генетическая изменчивость представляет собой генетический шум, обусловленный случайностью процесса мутагенеза, то у разных видов должны встречаться разные случайные наборы полиморфных локусов даже при выборе стандартного набора исследуемых генных локусов. Можно предположить, что гетерозиготности  $H$  всей кодирующей белки части генома у различных видов млекопитающих примерно одинаковы и характеризуются малым разбросом относительно среднего значения  $\bar{H}$  для всего класса. В этом случае числа полиморфных локусов, наблюдаемые в наборах из  $L$  локусов у *различных видов*, будут распределены по Пуассону так же, как и числа полиморфных локусов в наборах из  $L$  локусов, выбранных *в различных участках генома* одного и того же вида.

Автором собраны опубликованные данные по гетерозиготностям групп локусов  $H_L$  для 343 видов млекопитающих, рис. 1. Покажем, что полученное распределение частот встречаемости  $H_L$ , рис. 2, согласуется с распределением Пуассона.

Для каждого полиморфного локуса относительная частота  $\alpha$  его встречаемости в гетерозиготном состоянии определяется выражением:

$$\alpha = 1 - \sum_i q_i^2, \text{ где } q_i - \text{частоты встречаемости аллелей данного локуса.}$$

Предположим, что все  $l_p$  полиморфных локусов, дающих основной вклад в гетерозиготность  $H_L$  популяции, имеют одинаковые значения  $\alpha$ .

Тогда у каждой особи будет в среднем  $\alpha l_p$  гетерозиготных локусов. Средняя гетерозиготность особей популяции будет равна:

$$H_L = \alpha l_p / L . \quad (1)$$

При большом числе аллелей величина  $\alpha$  может быть сколь угодно близка к единице,  $\alpha_{max} \approx 1$ .

Для исследованного массива данных (рис. 1) среднее число измеряемых локусов  $\bar{L} = 26,7$ , средняя гетерозиготность  $\bar{H}_L = 0,041$ . Наименьшее число полиморфных локусов, которое может обеспечить заданную величину  $\bar{H}_L$ , равно  $l_{p\ min} = \bar{H}_L \bar{L} / \alpha_{max} = 1,1$ .

У млекопитающих полиморфны в среднем 20% генных локусов [3], т.е. в выборке из  $\bar{L} = 26,7$  локусов в среднем наблюдается 5,3 полиморфных локуса. Таким образом, аппроксимацию наблюдаемого распределения гетерозиготности  $H_L$ , рис. 2, следует искать в виде распределения Пуассона со средним числом полиморфных локусов  $\bar{l}_p = \bar{H}_L \bar{L} / \alpha$ , заключенным между  $l_{p\ min} \approx 1$  и  $l_{p\ max} \approx 5$ .

Процедура поиска наилучшей аппроксимации заключалась в следующем. Выбирались значения средней множественности Пуассона  $\bar{l}_{pi}$  от 1,0 до 3,3 через интервал  $\Delta \bar{l}_p = 0,1$  (в области максимальной значимости рис. 3 через интервалы  $\Delta \bar{l}_p = 0,5$ ) — всего 28 значений. Один полиморфный локус соответствует гетерозиготности  $\Delta \bar{H}_{Li} \equiv \bar{H}_L / \bar{l}_{pi}$ . Наблюдаемое распределение  $H_L$ , рис. 2, разбивалось на интервалы длиной  $\Delta \bar{H}_{Li}$ . Значениям гетерозиготности, меньшим  $0,5 \cdot \Delta \bar{H}_{Li}$  (0,5 полиморфного локуса) присваивалось целое значение 0 (0 полиморфных локусов). Значениям гетерозиготности от  $0,5 \cdot \Delta \bar{H}_{Li}$  до  $1,5 \cdot \Delta \bar{H}_{Li}$  (от 0,5 до 1,5 полиморфных локуса) — целое значение 1 (1 полиморфный локус) и т. д. Таким образом наблюдаемое непрерывное распределение гетерозиготностей  $H_L$  преобразовывалось в дискретное распределение натуральных чисел, которое сравнивалось с распределением Пуассона со средней множественностью  $\bar{l}_{pi}$ . При этом

для каждого  $\bar{l}_{pi}$  ( $i$  от 1 до 28) определялся уровень значимости аппроксимации по критерию согласия  $\chi^2$ , рис. 3.

Наилучшее (с наибольшими уровнями значимости  $S$  по критерию  $\chi^2$ ) согласие наблюдаемого распределения, рис. 2, с искомым распределением Пуассона получаем в области  $1,8 \leq \bar{l}_p \leq 2,2$  с максимумом  $S = 0,75$  при  $\bar{l}_p = 1,85$  и  $\alpha = \bar{H}_L \bar{L} / \bar{l}_p = 0,59$ , рис. 2 и 3. Как известно, распределение Пуассона со средней множественностью порядка единицы имеет максимум в нуле. Полученное малое значение средней множественности  $\bar{l}_p = 1,85$  позволяет объяснить высокую частоту обнаружения у различных видов нулевых значений  $H_L$ , рис. 1, которая часто вызывает беспокойство исследователей, рассматривающих генетическую изменчивость как адаптивный потенциал вида [6,7].

Покажем, что использование в (1) приближения, состоящего в предположении о том, что полиморфные локусы, дающие основной вклад в гетерозиготность  $H_L$  популяции, имеют одинаковые значения  $\alpha$ , является оправданным. В общем случае средняя гетерозиготность  $H_L$  популяции определяется присутствием в популяции полиморфных локусов с разными величинами  $\alpha$ . Непрерывность значений  $\alpha$  приводит к непрерывности наблюдаемого распределения гетерозиготностей. Разбивая всю область изменения величины  $\alpha$  на интервалы, в которые попадают значения  $\alpha$ , изменяющиеся в заданных пределах, можно представить величину гетерозиготности в виде суммы вкладов всех интервалов:

$$H_L = \left( \sum_i \alpha_i l_{pi} \right) / L \quad (2)$$

где  $\alpha_i$  – среднее значение  $\alpha$  в интервале  $i$ ;  $l_{pi}$  — число полиморфных локусов в выборке из  $L$  локусов со значениями  $\alpha$ , попадающими в  $i$ -й интервал. В силу случайности появления полиморфных локусов в

мутационном процессе числа их встречаемости  $l_{pi}$  в различных выборках из  $L$  локусов должны быть распределены по Пуассону. Гетерозиготность  $H_L$ , умноженная на соответствующий коэффициент, также должна быть распределена по Пуассону в тех случаях, когда главный вклад в сумму (2) дают один или несколько членов с примерно совпадающими значениями  $\alpha_i$ . Проведенный выше анализ доказывает, что сумма (2) обладает этим свойством. Независимый анализ данных рис. 1 с известными коэффициентами  $\alpha$ , проведенный автором, подтверждает это заключение. Оказалось, что для  $L \approx 27$  главный вклад в сумму (2) дают в среднем два полиморфных локуса с близкими величинами  $\alpha$  порядка 0,5. Суммарный вклад в сумму (2) локусов с малыми  $\alpha$  пренебрежимо мал.

Покажем теперь, что с увеличением числа измеряемых локусов  $L$  происходит сужение наблюдаемого распределения в соответствии со свойствами распределения Пуассона, т.е. с увеличением числа исследованных локусов  $L$  должна уменьшаться дисперсия наблюдаемого распределения гетерозиготности  $H_L$ .

Распределение Пуассона со средней множественностью  $\bar{l}_p$  и дисперсией  $\sigma_p^2 \equiv \overline{(l_p - \bar{l}_p)^2}$  характеризуется известным соотношением

$$\sigma_p^2 = \bar{l}_p. \quad (3)$$

Используя равенство  $\bar{l}_p = \bar{H}_L \bar{L} / \alpha$  и обозначение  $\sigma_{H_L}^2 \equiv \overline{(H_L - \bar{H}_L)^2}$ , можно переписать формулу (3) в виде:

$$\bar{l}_p = \frac{\bar{l}_p^2}{\sigma_p^2} = \frac{\bar{H}_L^2}{\sigma_{H_L}^2}, \quad \sigma_{H_L}^2 = \alpha \bar{H}_L / \bar{L}, \quad (4)$$

где  $\bar{H}_L = 0,041$ ,  $\alpha = 0,59$ . В общем случае при меняющемся  $\bar{L}$  можно предполагать аллометрическую зависимость  $\sigma_{H_L}^2$  от  $\bar{H}_L / \bar{L}$ , т.е. линейную связь логарифмов этих величин [8]:

$$\lg \sigma_{H_L}^2 = a \lg (\bar{H}_L / \bar{L}) + b. \quad (5)$$

Для распределения Пуассона в соответствии с (4) должны выполняться следующие равенства:

$$a = 1, \quad b = \lg \alpha = -0,23. \quad (6)$$

На рис. 4 проверены соотношения (5) и (6) для восьми интервалов изменения числа локусов  $L$  (см. подпись к рис. 1 для определения интервалов). Переменная  $\bar{L}$  в (4) принимала значения  $\bar{L}_i$  ( $i$  от I до VIII). Для каждого  $\bar{L}_i$  переменная  $H_L$  в (4) принимала значения, попадающие в  $i$ -й интервал (рис. 1). Значения  $a$  и  $b$  в линейной регрессии (5) (рис. 4) оказались равны

$$a = 1,1 \pm 0,2, \quad b = 0,1 \pm 0,7 \quad (r = 0,88, \quad P = 0,004)$$

Полученная оценка параметра  $b$  характеризуется большой погрешностью, что не позволяет определить значение  $b$  с достаточной точностью. В пределах этой погрешности предполагаемое значение  $b$  (6) не противоречит полученной оценке. Полученная величина  $a$  находится в хорошем согласии с соответственным соотношением (6).

Отметим, что если бы величина  $H$  была бы разной у разных видов, т.е. сильно варьировала бы относительно среднего значения  $\bar{H}$ , то при увеличении исследованного числа локусов  $L$  не происходило бы наблюдаемого сужения распределения  $H_L$  (рис. 4), так как в этом случае ширина распределения  $H_L$  определялась бы естественной шириной распределения  $H$ , которая от  $L$  не зависит. Отметим также, что в общем случае произвольного распределения  $H$  с большой дисперсией, распределение  $H_L$  не могло бы быть аппроксимировано распределением Пуассона (рис. 2), которое определяется случайными частотами встречаемости полиморфных локусов у разных видов при одинаковом среднем значении этих частот у всех видов.

Таким образом, проделанный анализ доказывает случайность распределения гетерозиготных локусов по геному млекопитающих и указывает на то, что генетическая изменчивость представляет собой

информационный шум. Он указывает также на то, что значения гетерозиготности  $H$  всей кодирующей белки части генома примерно одинаковы у различных видов млекопитающих. Они близки к наблюдаемому среднему значению  $\bar{H}_L$  гетерозиготности класса млекопитающих, полученному при усреднении по всем видам млекопитающих значений гетерозиготности  $H_L$ , наблюдаемых при исследовании небольших групп  $L$  локусов,  $\bar{H} = \bar{H}_L$ . Это позволяет интерпретировать  $\bar{H}_L$  как экологический порог допустимых значений гетерозиготности для класса млекопитающих.

Автор благодарит В. М. Гаврилова, Т. Б. Голубеву, В. Г. Горшкова и М. В. Филатова за многочисленные обсуждения.

#### СПИСОК ЛИТЕРАТУРЫ

1. Ayala F.J., Fitch W.M. // Proc. Nat. Acad. Sci. 1997. V. 94. P. 7691-7697.
2. Cook L.M. // Lecture notes in biomathematics. 1984. V. 53. P. 1-13.
3. Nevo E., Beiles A., Ben-Shlomo R. // Lecture notes in biomathematics. 1984. V. 53. P. 13-213.
4. Nei M. // 1984. Lecture notes in biomathematics. V. 53. P. 214-241.
5. Горшков В.Г., Макарьева А.М. // ДАН. 1997. Т. 332. '6. С. 802-806.
6. Merola M. // Conserv. Biol. 1994. V. 8. P. 961-971.
7. O'Brien S.J. // Conserv. Biol. 1994. V. 8. P. 1153-1155.
8. Peters R.H. The ecological implications of body size. Cambridge: Cambridge Univ. Press. 1983. 330 p.



Экологический порог внутривидовой генетической изменчивости  
млекопитающих

А. М. Макарьева

Наблюдаемая гетерозиготность видов млекопитающих обычно измеряется при исследовании небольшого числа (порядка двух-трех десятков) генных локусов генома. Собранные опубликованные значения гетерозиготностей изменяются от нуля до 18% у 343 естественных видов млекопитающих со средним значением для всего класса, равным 4,1%. В настоящей работе показано, что наблюдаемые гетерозиготности случайно выбираемых групп локусов не отражают величины истинных гетерозиготностей всей кодирующей белки части генома видов млекопитающих. Большой разброс наблюдаемых значений гетерозиготности связан с малостью числа исследуемых генных локусов. Показано, что с ростом числа исследуемых локусов разброс значений гетерозиготностей уменьшается в соответствии со свойствами распределения Пуассона, задаваемого случайным распределением полиморфных локусов по различным участкам генома. Отсюда следует, что наблюдаемое среднее значение гетерозиготностей, измеренных на небольших группах локусов у разных видов (4,1%), равно среднему значению гетерозиготности всей кодирующей белки части генома для класса млекопитающих. Истинные гетерозиготности всей кодирующей белки части генома млекопитающих примерно одинаковы у всех видов и характеризуются малым разбросом относительно среднего значения для всего класса. Это значение может рассматриваться как экологический порог гетерозиготности естественных видов млекопитающих, определяемый стабилизирующим отбором в условиях естественной экологической ниши.

## Summary

### Ecological threshold of intraspecific genetic variability in mammals

Makarieva A. M.

Protein heterozygosity in mammals is usually measured using small (20-30) numbers of gene loci. The published values of protein heterozygosity of 343 species of mammals cover an interval from 0 up to 18%, with an average of 4,1%. In this work it is shown that values of heterozygosity measured using small numbers of gene loci do not correlate with the true heterozygosity of the whole protein-coding region of mammalian genome. The wide scatter of the published data on protein heterozygosity is caused by the small number of loci studied. It is shown that with increasing number of studied loci the dispersion of heterozygosity values decreases in accordance with Poisson distribution, which is determined by random distribution of polymorphic loci over mammalian genome. This means that the mean value of heterozygosity measured in different species using small numbers of loci (4,1%) is equal to the true heterozygosity of the whole protein-coding region of mammalian genome. The latter value appears to be nearly the same in all species and is characterised by a narrow spread about the mean. This mean value of heterozygosity can be viewed as the ecological threshold of protein heterozygosity in natural species of mammals, which is determined by stabilising selection in the natural ecological niche of mammals.

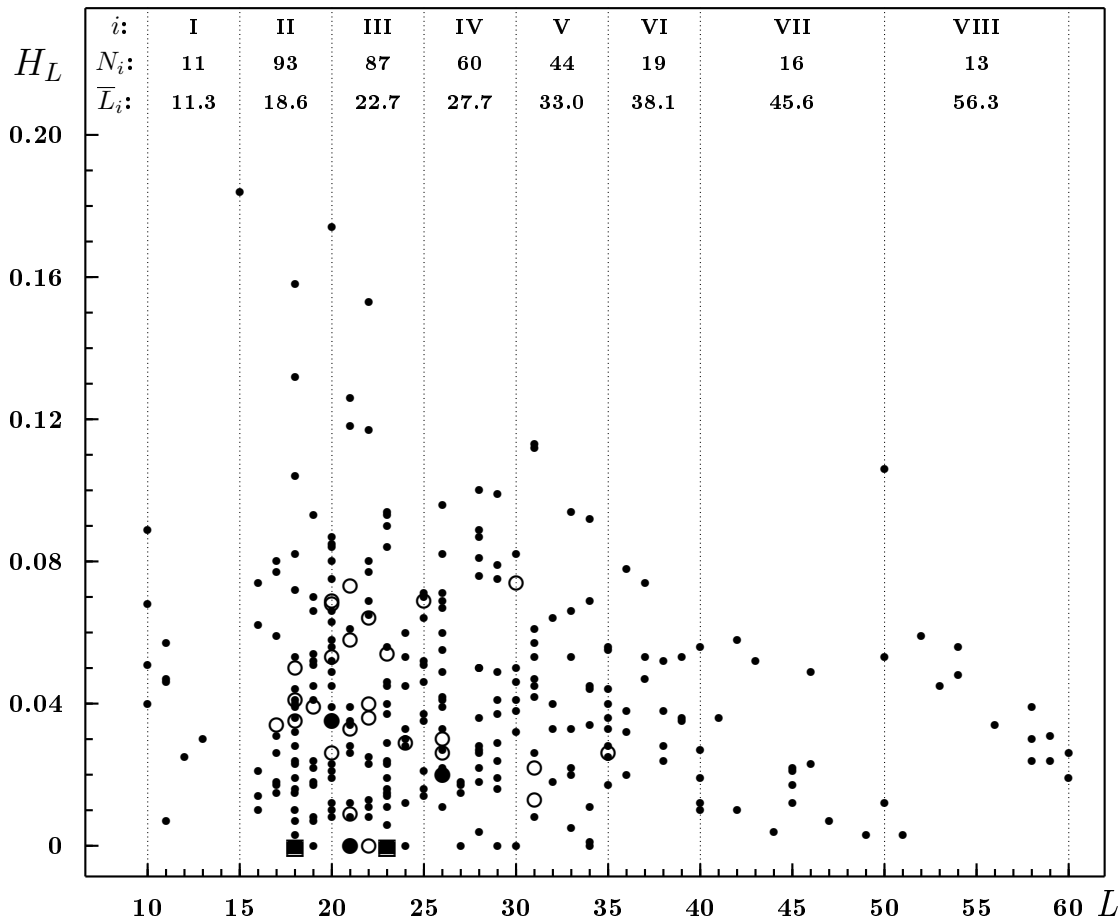


Рис. 1. Гетерозиготности группы локусов  $H_L$  различных естественных видов млекопитающих в зависимости от числа исследованных локусов  $L$ .

Точки – единичные значения; крупные полые кружки, крупные зачерненные кружки и крупные зачерненные квадраты обозначают 2, 3 и 4 случайно совпавших значения измерений для разных видов, соответственно. Пунктирные линии разбивают весь массив данных на восемь интервалов так, чтобы количество значений в каждом интервале было не менее 10 (этим объясняется бóльшая ширина интервалов VII и VIII). Значения, попадающие на границы интервалов, относились к левому интервалу, за исключением значений с  $L = 10$ . Вверху даны количества значений  $N_i$ , попадающих в  $i$ -й интервал, и средние значения  $\bar{L}_i$  числа локусов для каждого интервала.

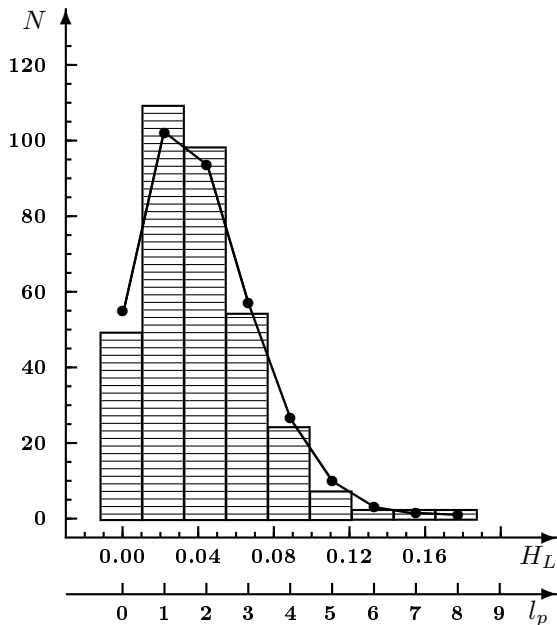


Рис. 2. Распределение частот встречаемости гетерозиготностей млекопитающих  $H_L$ . Заштрихованная гистограмма – наблюдаемые значения, построенные по данным рис. 1. Точки, соединенные сплошной линией, – наилучшая (с максимальным уровнем значимости  $S = 0,75$  по критерию  $\chi^2$ , рис. 3) аппроксимация наблюдаемого распределения гетерозиготности  $H_L$  распределением Пуассона для чисел полиморфных локусов  $l_p$  (нижняя шкала), линейно связанных с гетерозиготностью  $H_L$  (см. текст). Площадь под гистограммой равна числу исследованных видов.

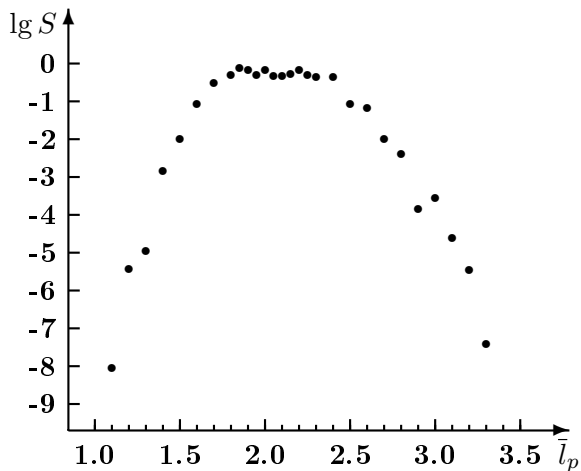


Рис. 3. Логарифмическая зависимость уровня значимости  $S$  по критерию  $\chi^2$  аппроксимации наблюдаемого распределения гетерозиготности  $H_L$  распределением Пуассона от величины средней множественности распределения Пуассона  $\bar{l}_p$ . Процедуру преобразования непрерывного распределения  $H_L$  в дискретное распределение натуральных чисел см. в тексте.

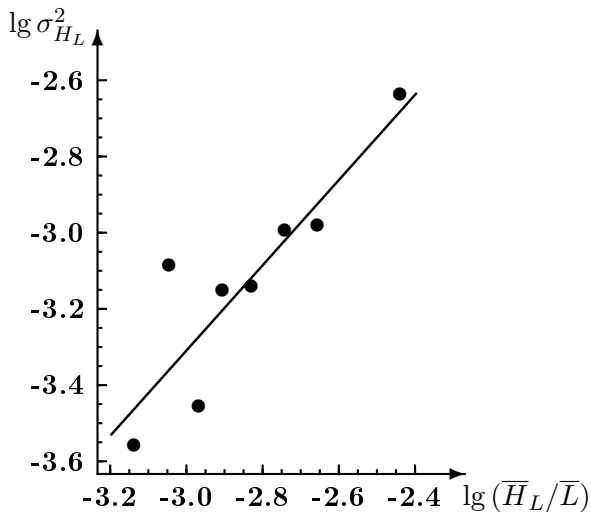


Рис. 4. Зависимость дисперсии  $\sigma_{HL}^2 \equiv \overline{(H_L - \overline{H}_L)^2}$  гетерозиготности  $H_L$  млекопитающих от среднего числа исследуемых локусов  $\overline{L}$  в форме аллометрической связи (4). Параметры линейной регрессии приведены в тексте.